

1. (Twice Amended) A cache memory for three-dimensional graphics texture mapping, comprising:

sub-pu
C1
a first DRAM bank for storing all texels of predetermined upper levels of LOD (Level of Detail) and a second DRAM bank for storing only a working set currently needed among the remaining levels of LOD, wherein said first and second DRAM banks including SAM ports, respectively, each of said SAM ports reading a texture for a trilinear interpolation and fetching new texture sub-clips from the outside;

a sub-clip loader connected to said SAM ports of said first and second DRAM banks and for fetching new texture sub-clips from an external system memory;

a sub-clip predictor for performing a hardware-based prefetch of a sub-clip to be soon needed,

a controller for controlling said components; and

a CAM for checking if eight texels existing at an integer coordinate relative to an LOD and (u, v) coordinates are located in said first and second DRAM banks, when the LOD and (u, v) coordinates mapped into a texture space with respect to a pixel to be rendered on a display screen are input to said controller.

5. (Twice Amended) The cache memory according to claim 1, wherein said sub-clip predictor prefetch sub-clip predicted to be soon needed in a 3-dimensional space, so as to reduce a penalty due to cache miss.

6. (Amended) A method for reducing a penalty occurring upon a cache miss, comprising the steps of:

performing a sub-clip prediction in one stack layer, where, under a hardware-based sub-clip prediction limit 2 by 2 sub-clip boundary inside of sub-clips (4 by 4) one a current clip RAM stack, when the tracing of (u, v) coordinates passes the limit, sub-clips of the tracing direction (left, right, upper and lower sides) are prefetched; and

performing a stack layer prediction in 3-dimensional space, where the current clip RAM stack represents the levels of LOD (LOD I to LOD i+3_ stored in said stack, internal two levels of LOD are used as a prediction limit, and, immediately after the tracing of LOD passes the limit, a stack layer corresponding to a next level of LOD is prefetched.